DOCUMENT RESUME

ED 126 572 EA 008 514

TITLE Three Strategies for Studying the Effects of School Processes.

INSTITUTION New York State Education Dept., Albany. Bureau of School Programs Evaluation.

PUB DATE Mar 76

NOTE 62p.; An expanded edition of "Which School Factors Relate to Learning?"

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.

DESCRIPTORS *Academic Achievement; Classroom Observation Techniques; *Educational Research; Elementary Education; *Performance Factors; Statistical Analysis; *Tables (Data)

IDENTIFIERS *New York

ABSTRACT

This report summarizes three recent studies conducted by the Bureau of School Programs Evaluation of the New York State Education Department that sought to identify relationships between various school factors and student performance in reading and mathematics. The studies utilized three different research strategies. The Regression Study statistically analyzed State Education Department data to determine how various school factors relate to student achievement. The Outlier Study identified high- and low-performing schools through statistical analysis and then analyzed available data for those schools in an attempt to find distinguishing variables. The Observational Study involved observing in 14 of the high- and low-performing schools identified in the Outlier Study in an attempt to discover classroom processes that differentiated between the contrasting groups of schools. Part 1 of the report is a nontechnical summary of the findings and conclusions of all three studies. Parts 2, 3, and 4 provide more detailed descriptions of each of the three studies individually. (Author/JG)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED126572

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

THREE STRATEGIES FOR STUDYING THE EFFECTS OF SCHOOL PROCESSES

An Expanded Edition of
Which School Factors Relate to Learning?

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Bureau of School Programs Evaluation
Albany, New York 12234

March 1976

EA 008 514

THE UNIVERSITY OF THE STATE OF NEW YORK

Regents of The University (with years when terms expire)

1981	Theodore M. Black, A.B., Litt.D., LL.D., Pd.D., D.C.L. Chancellor	Sands Point
1987	Carl H. Pforzheimer, Jr., A.B., M.B.A., D.C.S., H.H.D. Vice Chancellor	Purchase
1978	Alexander J. Allan, Jr., LL.D., Litt.D.	Troy
1980	Joseph T. King, LL.B.	Shelter Island
1981	Joseph C. Indelicato, M.D., L.H.D.	Brooklyn
1979	Francis W. McGinley, B.S., J.D., LL.D.	Glens Falls
1986	Kenneth B. Clark, A.B., M.S., Ph.D., LL.D., L.H.D., D.Sc.	Hastings on Hudson
1983	Harold E. Newcomb, B.A.	Owego
1988	Willard A. Genrich, LL.B., L.H.D., LL.D.	Buffalo
1982	Emlyn I. Griffith, A.B., J.D.	Rome
1977	Genevieve S. Klein, B.S., M.A.	Bayside
1981	William Jovanovich, A.B., LL.D., Litt.D., L.H.D.	Briarcliff Manor
1976	Mary Alice Kendall, B.S.	Irondequoit
1984	Jorge L. Batista, B.A., J.D.	Bronx
1982	Louis E. Yavner, LL.B.	New York

President of The University and Commissioner of Education
Ewald B. Nyquist

Executive Deputy Commissioner of Education
Gordon M. Ambach

Deputy Commissioner for Elementary, Secondary, and Continuing Education
Thomas D. Sheldon

Associate Commissioner for Research, Planning and Evaluation
John W. Polley

Director, Division of Research and Evaluation
Carl E. Wedekind

Director, Division of Evaluation
Alan G. Robertson

Chief, Bureau of School Programs Evaluation
David J. Irvine

FOREWORD

The Bureau of School Programs Evaluation of the New York State Education Department has for a number of years been developing procedures for evaluating the performance of schools and school districts. The Quality Measurement Project, which began in the late 1950's, assessed the academic achievement of school districts in relation to students' general ability, parents' education, and fathers' occupations. The Performance Indicators in Education project, starting a decade later, used other socioeconomic factors in the community to develop expected achievement levels for school districts against which the districts' actual achievement levels could be compared. Thus, a district's performance was based, not on how the district compared with other districts, but on how it compared with a unique standard derived from its own characteristics.

Paralleling this interest in the output of school districts has been a concern for the processes operating within school systems. To summarize research findings on this topic, a review was made of almost 100 studies dealing with factors related to student performance. Two reports resulted: one, a detailed description of methods and results;[1] the other, a summary which related research findings to ten important questions about the effectiveness of education.[2]

With this background, the bureau launched its own studies of school processes to learn more about what schools can do to improve the achievement of students. Three basic strategies were used. One strategy involved carrying out statistical analyses of data available in the Department's data files to determine how various school factors relate to student achievement. Both the second and third strategies involved identifying high- and low-performing schools. Then, under the second strategy, available data for the schools were analyzed to find variables which distinguished the two groups. The third strategy called for observing in high- and low-performing schools to discover classroom processes which differentiated between the contrasting groups of schools.

The present report is a summary of these studies. The studies have a common purpose: to identify relationships between school factors and student performance. While the reader is urged to be cautious in inferring cause and effect from these relationships, it is hoped that the information resulting from the studies will be useful in setting policy and making decisions.

This document is a result of efforts of a number of individuals. The Outlier Study, described in Part III, was carried out by Austin D. Swanson, Professor of Educational Administration, and Robert C. Nichols, Professor of Educational Psychology, both of the University of New York at

Buffalo. The Observational Study, described in Part IV, was carried out by Richard M. Clark, Professor of Educational Psychology of the University of New York at Albany. The Regression Studies were carried out by the staff of the Bureau of School Programs Evaluation: David J. Irvine, Chief; Gerald H. Wohlferd; Guy D. Spath; and Philip J. Pillsworth. Gennaro DiGiovanni, who served as a public administration intern during the time these studies were under way, conducted several of the regression analyses. Mr. Spath coordinated the consolidation of the several studies into this report.

John W. Polley
Associate Commissioner for
Research, Planning, and
Evaluation

TABLE OF CONTENTS

Part		Page
	FOREWORD	iii
	INTRODUCTION	1
I	AN OVERVIEW OF THREE SETS OF STUDIES	2
	THE STUDIES	3
	Regression Studies	3
	The Outlier Study	3
	The Observational Study	4
	FINDINGS	4
	Is Size a Factor?	4
	Do Expenditures Affect Learning?	5
	How Are Special Programs and Services Related to Achievement?	5
	Traditional vs. Open Classrooms	6
	Do Teacher Characteristics Make a Difference?	6
	How Do School Conditions and Practices Influence Learning?	8
	How Do the Attitudes of the Staff Relate to Learning?	8
	SUMMARY OF FINDINGS	9
	Factors Associated with High Achievement	10
	Factors Associated with Low Achievement	11
	Factors Showing Ambiguous Relationships with Achievement	11
	Factors Not Associated with Achievement	11
	DISCUSSION	12
II	THE REGRESSION STUDIES	13
	Studying School District Processes	14
	Attendance Rate	17
	Mobility	17
	Enrollment	17
	Teacher Characteristics	17
	Teacher Salaries	18
	Conclusion	19

Part		Page
III	THE OUTLIER STUDY	20
	A Strategy for Studying School Processes	21
	Determining Performance of Schools	22
	Findings	23
	Conclusions	28
IV	THE OBSERVATIONAL STUDY	30
	Background	31
	Instrument Development	32
	Procedures for School Visits	34
	Analysis of Data: Overview	34
	General Classroom Observations	35
	Teacher Questionnaire	37
	Principal Interview	39
	Observation of Reading	40
	Reading Teacher Interview	42
	Teacher Reinforcement Scale	43
	Characteristics of Open Education	44
	Discrimination Between High- and Low-Performing Schools	46
	Observer Agreement	48
	Discussion	49
	Summary of Most Significant Findings	50
	REFERENCES	52

LIST OF TABLES

Table		Page
1	Means and Standard Deviations of Criterion and Control Variables for Six Groups of Deviate and Non-deviate Schools	24
2	Means and Levels of Significance of School and Program Variables Which Differentiate Among the Groups with a Chance Probability of Less Than .05	26
3	Means and Levels of Significance of Achievement-Related Variables for 1971-72 Which Differentiate Among Groups with a Chance Probability of Less Than .05	27
4	Means and Levels of Significance of Variables Describing Characteristics of Teaching Personnel Which Differentiate Among the Groups with a Chance Probability of Less Than .05 for Public Schools Only	29
5	Means for High- and Low-Performing Schools on General Classroom Observation Form	36
6	Mean Expectation for Pupils Expressed by Teachers in High- and Low-Performing Schools	37
7	Mean Perception of Pupils Expressed by Teachers in High- and Low-Performing Schools	38
8	Mean Expectation for Support Expressed by Teachers in High- and Low-Performing Schools	38
9	Principal Ratings of Personnel as Ascertained from Interviews	39
10	Principal Ratings of Rapport as Ascertained from Interviews	40
11	Mean Ratings of Activities Carried on in Reading Classes in High- and Low-Performing Schools	41
12	Means of Ratings of Activities of Non-Reading Groups During Reading Period	42

Table		Page
13	Mean Responses of Reading Teachers in High- and Low-Performing Schools	43
14	Mean Scores for Teacher Reinforcement in Classes in High- and Low-Performing Schools	44
15	Comparison of Mean Scores of Characteristics of Open Education in High- and Low-Performing Schools	45
16	Correlations of Selected Items Related to Open Education in Grades 1-3	46
17	Discrimination of Rank-orders of High- and Low-Performing Schools on Selected Items on Open Education	47
18	Discrimination of Rank-orders of High- and Low-Performing Schools on Selected Items on General Classroom Observation	47
19	Discrimination of Rank-orders of High- and Low-Performing Schools on Selected Items on Observation of Reading	48
20	Root Mean Square Differences Between Observers	49

LIST OF FIGURES

Figure		
1	1970 Percent Uncertified Teachers and 1971 Third-Grade Mathematics with Socioeconomic Factors Controlled	16
2	Relationship of Median District Teacher Salary to Reading and Mathematics Achievement with Socioeconomic Factors Controlled	18

INTRODUCTION

Public interest in education has increased in recent years. The causes of this increased interest are diverse. Education accounts for the single largest government expenditure in New York State, to cite one possible cause; and in time of economic pressures, taxpayers and government officials scrutinize expanding budgets. Beyond fiscal concerns, the basic philosophy of American education is being questioned. Education has long been considered one of the most important routes to success. Disillusionment with massive attempts to improve education has brought this belief under fire. Finally, much educational research has emphasized the importance of student background and, by contrast, de-emphasized the importance of the schools in stimulating students' learning. This has led many people to ask, "Does schooling make a difference?"

It is not sufficient to say that schooling does or does not do what is expected of it. In order to improve the schools it is necessary to understand the effects of various school factors on different students. Many effects are logical or are known intuitively to educators and the public. But the difficulty of improving the performance of students, even with massive infusions of funds, has illustrated the limits of our knowledge. For this reason, educational research has turned more and more to attempting to discover which school factors relate to student learning. This is the direction explored by the research reported in the present document.

Some cautions are in order. This document focuses on school factors as they relate to student performance in reading and mathematics only. It is possible that these school factors relate to other kinds of student performance in different ways. It is also possible that factors found here to be unrelated to student performance in reading and mathematics are quite important for other kinds of performance. Still other factors may be important to the quality of life in educational institutions although they have no apparent impact on student performance. For these reasons, changes made in school programs to improve student performance in reading and mathematics should be examined for possible effects on other aspects of student functioning.

This report is a summary of research conducted by the staff of the Bureau of School Programs Evaluation and by consultants under contract to the bureau. The report is divided into four parts. Part I is a nontechnical summary of findings and conclusions of all the studies. It can be read independently of the other three parts. Parts II, III, and IV are more detailed descriptions of the three types of studies conducted.

PART I

AN OVERVIEW OF THREE SETS OF STUDIES

PART I

AN OVERVIEW OF THREE SETS OF STUDIES

Three sets of studies of school factors were conducted under the aegis of the Bureau of School Programs Evaluation. All of the studies had the purpose of identifying school factors which relate to the achievement of students. Part I of this report describes the three sets of studies and summarizes the findings.

This report was designed to complement an earlier bureau publication, What Research Says About Improving Student Performance.[2] That publication summarized the results of almost 100 research studies dealing with the relation of a variety of school factors to school outcomes. The present report attempts to integrate results of research done more recently in the schools of New York State.

It is hoped that the results reported here can provide a basis for thoughtful discussion and, together with other information, suggest directions which can be pursued to improve education. By its very nature, the research does not offer simple answers to the complex questions confronting education today. But if the findings seem to be cautiously interpreted, perhaps that fact will help us avoid rushing to conclusions or jumping on bandwagons.

THE STUDIES

Regression Studies

The first set of studies, carried out by the bureau staff, investigated factors which are at least partly under the control of school personnel. As in the Coleman Report, the effects of nonschool factors were controlled for and the unique contributions of school variables were studied. The studies as a group are referred to as the Regression Studies and are described in more detail in Part II.

The Outlier Study

The second type of study used multiple regression analysis to identify schools which were performing either above or below their predicted levels of achievement, as computed from nonschool factors. The study focused on school buildings rather than school districts. Three groups of schools were identified: 1) High outliers, those schools whose actual mean achievement scores were well above their predicted scores; 2) Low outliers, those schools whose actual scores were well below their predicted scores; and 3) Midliers, those schools whose actual scores fell near their predicted scores. These three groups of

schools were compared through analysis of variance in an attempt to determine the effects of a number of variables which reflect school processes. Because this study emphasized schools which lay some distance from their expected levels of achievement, it was dubbed the Outlier Study. A more complete description can be found in Part III.

The Observational Study

The third kind of study involved observing in 14 schools which were identified as above or below predicted achievement in the Outlier Study described above. Observational instruments, interviews, and questionnaires were used to obtain information about classroom activities, interactions among students and staff members, and perceptions of staff members. This study is referred to as the Observational Study and is described more fully in Part IV.

FINDINGS

While each of the sets of studies described above used a somewhat different approach, their findings are consolidated below in an attempt to develop a coherent, though not necessarily complete, picture of how school processes relate to school outcomes. Where the results are contradictory or ambiguous, an attempt is made to show this.

Is Size a Factor?

The average school district enrollment in New York State in 1971 and 1972 was approximately 2500, ranging from 30 to over 30,000. In the Regression Studies, district enrollment was found to be negatively related to achievement; that is, larger districts had poorer average achievement. However, when total population of districts was considered, it replaced enrollment as a predictor. This seems to suggest that the negative relationship between enrollment and achievement is a function of urbanness rather than of school size. The finding, then, does not appear to offer evidence about the optimum size of schools or school districts.

The Outlier Study showed no difference in size between high and low outlier schools; each group averaged about 100 fewer students than did schools identified as midliers. In addition, classes tended to be smaller for both positive and negative outliers than for midliers. As has been suggested in other studies, the effect of class size may be dependent on the type of student and on the subject being taught.

Do Expenditures Affect Learning?

Several types of expenditures were investigated. Since the largest single expenditure for education is for salaries, teachers' salaries were studied in several different ways.

The Regression Studies revealed a positive relationship between median teacher salaries of school districts and average student achievement in the districts. These findings were supported by the Outliers Study, in which teachers in schools achieving above expected earned the most money and teachers in schools achieving below expected received the least. While these findings do not mean that higher salaries produce higher performance, they indicate that the more successful districts do pay higher salaries. Not unexpectedly, salary-related variables, such as graduate credits and experience of teachers, were also shown to be related to achievement. It is interesting to note that an earlier study, using 1969 data, showed that the amount of money spent per pupil on principals' salaries was also related to achievement.

The Regression Studies also examined full tax value and several district per-pupil expenditures, including total expenditures and expenditures for regular day instruction, for teachers, for central administration, and for principals. No consistent relationships were found, in spite of the fact that wide variations were observed from district to district.

How Are Special Programs and Services Related to Achievement?

Special programs frequently showed negative relationships with achievement. The Outlier Study, for example, revealed a negative relationship between achievement and special programs for the handicapped. Guidance, social work services, and attendance services were also negatively related to achievement, while programs for the academically talented were more likely to be found in the high outlier schools. These findings can be misleading if cause and effect are assumed. Such programs are most likely to be found where they are most needed. Therefore, it is not surprising to find that more special programs are found where achievement is low. In addition, the data available for these studies did not clearly distinguish between types of special programs; it is likely that some types of programs affect students differently from other types of programs.

In the Observational Study, high-achieving schools were rated as having higher total activity in nine of eleven reading activities, especially in silent reading, than did low-achieving schools. The Outlier Study showed that the use of rooms for academic rather than vocational study also correlated positively with achievement. The Regression Studies indicated a positive relation between attendance rate and achievement, but this relationship disappeared when socioeconomic factors were considered. The same phenomenon occurred in a study of student mobility.

Traditional vs. Open Classrooms

Attempts to move away from traditional, teacher-centered classroom arrangements have included many innovations designed to increase learning. The Outlier Study showed that most classrooms in the schools studied were traditional, with the smallest number of traditional classrooms being found in low outlier schools and the greatest number in midlier schools. High outliers were more likely to have open classrooms and multi-age groupings. Multi-unit plans were negatively related to performance. Midlier schools were less likely to have innovative programs than either of the other two groups of schools. Other organizational arrangements were not significantly related to achievement. These included clustering, continuous progress, departmentalization, differentiated staffing, dual progress plan, house plan, modular scheduling, non-graded, self-contained, and team teaching. The Observational Study supported the findings of the Outlier Study in respect to open classrooms.

Do Teacher Characteristics Make a Difference?

When candidates for teaching positions walk into a superintendent's office, they bring with them certain personal and professional characteristics. Among them are their professional training and experience, age, sex, marital status, and a variety of personal traits. The superintendent may have very little concrete evidence to use in selecting among candidates with an almost infinite number of combinations of professional and personal characteristics. Does a candidate with a doctorate have more to offer than one with many years of experience but less formal education? Are women more effective than men in elementary schools? How important is graduate school training when the teacher is expected to teach reading to eight-year-old children?

The New York State Education Department collects data on teachers in each of the school districts in the state. From the available data, five teacher characteristic variables were selected for study:

1. Median Age of Teachers in the District
2. Median Years of Experience
3. Percent of Married Teachers
4. Percent of Male Teachers
5. Percent of Teachers Having Graduate Credit

The Regression Studies revealed no relationships between teacher age or experience and average student achievement in a district. The Outliers Study, on the other hand, showed a positive relationship between teacher experience and student performance. Furthermore, a larger percentage of the teachers in the high outlier schools were on tenure.

The percent of teachers having some graduate training showed the most consistent relationship with reading and mathematics achievement in the Regression Studies. Even after controlling for socioeconomic factors, percent of teachers having 30 or more credit hours beyond the bachelor's degree demonstrated a significant relationship to all of the achievement criteria.

Significant shifts in teacher certification took place from 1970 to 1971. In 1970 only 230 districts had all teachers certified while in 1971 this number increased to over 480. The Regression Studies of this variable showed a definite negative relationship of percent uncertified to student achievement in 1970. The tremendous reduction in number of schools having uncertified teachers in 1971 led to less conclusive results for that year, but the inference still seems warranted that certification is desirable.

The positive findings on certification and graduate training were both supported in the school-building analyses carried out in the Outliers Study.

Using regression analysis to study school buildings, it was found that schools with a greater percentage of Black teachers had higher levels of performance, after controlling for non-school factors.

The findings on sex and marital status were mixed. In 1971, no relationship was found between percent of married teachers in a district and achievement. In 1972, a positive relationship was evident. Percent of male teachers was negatively related to all achievement criteria in 1972 but only to sixth-grade mathematics in 1971.

While these results suggest that relationships exist between student achievement and teacher graduate training, it should be noted that these relationships may reflect other factors related to both achievement and teacher characteristics. For instance, we know that low socioeconomic school districts tend to have low mean scores on achievement tests and high socioeconomic districts tend to have high mean scores. The positive relationship that seems to exist between percent of teachers with graduate credits and student achievement may simply reflect a tendency among higher socioeconomic districts to employ teachers with graduate training. Conversely, the lower socioeconomic districts may not have the money to pay the higher salaries of teachers with graduate credits.

The hypothesis that these teacher characteristic relationships are merely reflective of the known relationship between socioeconomic status and achievement is even more tenable regarding teacher marital status and sex. Lower socioeconomic districts might be expected to hire more men since women might be unwilling to teach in those districts. And, with a higher teacher turnover rate, these districts could be expected to have a lower percentage of married teachers.

How Do School Conditions and Practices Influence Learning?

Logic suggests that school conditions and practices which impinge directly on the daily lives of students offer the most promise for improving education. Yet in many ways they are the hardest to study.

The Observational Study attempted to obtain data which would make it possible to understand better the importance of certain conditions and practices. Some of the results are summarized below:

1. Teachers in high outlier schools made less overt effort to maintain class control, had less rigid student behavior, but were more efficient in maintaining the level of control they appeared to want than were teachers in low outlier schools.

2. Teachers in high outlier schools were rated as warmer, more responsive, and placing more emphasis on cognitive development.

3. More total activity takes place in reading classes in the high outlier schools.

4. Children in high outlier schools engage in more silent reading while children in low outlier schools engage in more oral reading.

5. In grades one to three, teachers in high outlier schools gave more positive and less negative reinforcement than did teachers in the low outlier schools.

6. In grades four to six, teachers in low outlier schools gave more reinforcement. In general, however, they tended to use negative reinforcement more than the teachers in the high outlier schools.

7. Pupils in the high outlier schools were more enthusiastic and were better able to sustain attention.

8. On selected items related to open education, the high outlier schools appeared more often.

9. Items on physical space and facilities generally did not differentiate between high and low outlier schools.

How Do the Attitudes of the Staff Relate to Learning?

The attitudes of teachers are frequently cited as influences on the performance of students. For that reason a number of attitudes, perceptions, and expectations of teachers and other staff members were

investigated in the Observational Study. The findings include the following:

1. Teachers in high outlier schools expected more children to graduate from high school, to go to college, to become good readers and to become good citizens than did teachers in low outlier schools.

2. Teachers in high performing schools saw their children as more intelligent, better behaved, more pleasant to teach, and their parents as more concerned.

3. Teachers in high and low outlier schools were not different in the amounts of help they perceived as being available in handling problems.

4. Reading teachers in the high outlier schools gave more favorable evaluations of the reading programs in their schools than did reading teachers in low outlier schools. They also rated the classroom teachers more favorably in using appropriate materials, extending reading into other areas, asking children to read with purpose, and using informal diagnosis.

5. Principals in high outlier schools generally saw their personnel as more competent than did principals in low outlier schools.

6. Principals in high outlier schools saw themselves as having better rapport with teachers, parents, and pupils than did the principals in low outlier schools. However, principals in low outlier schools reported better rapport with the school board.

SUMMARY OF FINDINGS

Summarizing the findings of a group of related studies presents a number of difficulties. The volume of results resists a concise treatment. Attempting to discuss the results in simple terms may produce misleading conclusions. The ambiguities and contradictions between the findings of different studies may defy easy explanations.

A particular problem in interpreting cross-sectional data, which these studies used, involves the extent to which variables can be inferred to cause the outcomes with which they are associated. It is part of the litany of research that "correlation does not imply causation." However, there is a human tendency to jump to conclusions about cause and effect. Findings such as those presented here should be interpreted with restraint and logic in order to avoid faulty conclusions.

One area in which a logical analysis of the situation may avoid incorrect conclusions has to do with the findings that special programs are frequently associated with low achievement. A hasty conclusion may be that the special programs are ineffective or actually detrimental to

student achievement. However, many special programs--for the handicapped or for the disadvantaged, to cite two instances--have been implemented to meet particular needs. A negative correlation between the prevalence of special programs and achievement, rather than meaning that the programs have adversely affected achievement, indicates that the programs are located where they are needed.

In spite of the possible problems of interpretation, it seems that a summary of findings from the three types of studies may be useful. The following summary shows which variables were associated with students' achievement in reading and arithmetic after social and economic factors were accounted for.

Factors Associated with High Achievement

High achievement of students was associated with the following factors:

1. Higher teachers' salaries.
2. Use of rooms for academic rather than vocational study.
3. Open classrooms.
5. Multi-age groupings.
5. Higher levels of graduate training of teachers.
6. Larger percent of Black teachers.
7. Better control of classes but with less overt effort on the part of teachers to maintain control.
8. Less rigid student behavior.
9. Greater teacher warmth and responsiveness.
10. Greater emphasis by teachers on cognitive development.
11. More total activity in reading classes.
12. More silent reading.
13. Positive reinforcement of students by teachers.
14. More enthusiasm on the part of students.
15. Better ability on the part of students to sustain attention.
16. Higher expectations on the part of teachers for their students to become good readers and good citizens, to graduate from high school, and to go to college.
17. Teachers' perceptions of their students as more intelligent, better behaved, and more pleasant to teach and the students' parents as more concerned.
18. More favorable ratings by reading teachers of the reading program in their schools.

19. More favorable ratings by reading teachers of classroom teachers in using appropriate materials, extending reading into other areas, asking children to read with purpose, and using informal diagnosis.
20. Higher ratings by principals of the competence of personnel in their schools.
21. Principals' perceptions of a high level of rapport with teachers, parents, and pupils.

Factors Associated with Low Achievement

Low achievement of students was associated with the following factors:

1. Larger district enrollment.
2. Special programs for the handicapped.
3. Pupil services, including guidance, social services, and attendance services.
4. Multi-unit groupings.
5. Larger percent of uncertified teachers in a district or school.
6. More oral reading.
7. Negative reinforcement of students by teachers.
8. Principals' perceptions of a high level of rapport with the school board.

Factors Showing Ambiguous Relationships with Achievement

A number of variables showed ambiguous relationships with student achievement. Among these were:

1. Class size.
2. Per-pupil expenditures for instruction, teachers, principals, and central administration as well as total per-pupil expenditures.
3. Median years of experience of teachers in a district.
4. Percent of married teachers in a district.
5. Percent of male teachers in a district.

Factors Not Associated with Achievement

Several variables were found not to be associated with achievement. Among these were:

1. Attendance rate, once socioeconomic factors are considered.
2. Student mobility.

3. Various organizational and grouping arrangements, including clustering, continuous progress, departmentalization, differentiated staffing, dual progress plan, house plan, modular scheduling, non-graded, self-contained, and team teaching.
4. Median age of teachers in a district.
5. School facilities and space.
6. Teachers' perceptions of the amount of help available in handling problems.

DISCUSSION

In many respects, the findings of these studies seem to agree with other research conducted during the past decade. However, some very logical, long-held assumptions about the effects of certain system-wide administrative variables are not supported. None of the "expected" relationships with achievement were demonstrated for attendance, mobility, special compensatory programs and services, and gross expenditure variables.

Mixed results were found for class size, school size, and innovative programs. Smaller classes in smaller schools and innovative programs appear to be found in both high and low outliers.

The most encouraging findings related to teacher characteristics, staff attitudes, and school practices. Results of the several studies indicate that good teachers are the heart of the educational system, as conventional wisdom would suggest. Students seem more likely to achieve well where teachers are better trained, more often certified, higher paid, and more likely to be tenured. The teachers in high performing schools have higher expectations for their students and more favorable perceptions of them; they appear to be warmer, more supportive, and more responsive. This is accompanied by more enthusiastic students. Teachers in more successful schools also appear to deemphasize strict control and lean toward more open education. These studies do not inform us as to what "better" teachers do to bring about high achievement, nor do they explain why a warmer, more open environment is more conducive to learning.

The relationships described are not necessarily causal. Generalizations drawn from these studies should be considered in the light of other research and the decision maker's experience and unique situation. These findings are offered as one more bit of information which can contribute to an understanding of educational processes. One conclusion seems apparent: Studying district-wide variables and school variables seems to be less rewarding than studying the teaching-learning interface.

PART II

THE REGRESSION STUDIES

PART II

THE REGRESSION STUDIES

The Performance Indicators in Education (PIE) project was initiated by the State Education Department with the purpose of developing new ways of studying the performance of schools and school districts. The major thrust has been toward reporting school district performance while taking into account the characteristics of the district and its students.

School district performance was estimated by analyzing data collected from various sources including the U. S. Census, the Basic Educational Data System, and the Pupil Evaluation Program (PEP). The primary statistical tool used in the project is multiple regression analysis. By analyzing socioeconomic data and achievement data for school districts, combinations of variables were identified which predict average district scores on reading and mathematics, as measured by the PEP tests. Each set of relationships was expressed as an equation, which included a weight for each of the variables contributing to the prediction of achievement. By substituting in the equation the values of the predictor variables for a given district, a predicted score was obtained for that district on that particular measure of achievement. The difference between the actual score for the district and this predicted score was its performance indicator.

Performance indicators provide a better estimate of school district performance than does the mean achievement test score for the district, since they represent factors outside the control of the schools which affect student performance. We know, for instance, that socioeconomic status correlates highly with school achievement. This suggests that a school which has a heavy concentration of students from high socioeconomic families is likely to have some educational advantages not shared by all districts. To compare such a district's test scores with those of districts with different characteristics is not very meaningful.

To summarize, the original intent of the Performance Indicators Project was to take into account each district's unique characteristics so that a more realistic picture of the district's effectiveness could be obtained. A more detailed description of the procedure can be found in the technical manual prepared for the 1974 report.[3]

Studying School District Processes

The models developed to assess district performance describe the status of a district's functioning in certain areas of its curriculum. In order to improve the performance of school districts, it is necessary to determine what causes the district to function as it does and to infer what changes will bring about improvement.

The performance indicators models served as a foundation for studying school processes. They were used to control statistically for the effects of nonschool factors in order to study the relationships of school conditions and resources to pupil achievement.

Variables used to control for nonschool factors included:

- Total Population of the District
- Percent Rural Population
- Percent of Children Living in a Father/Mother Family
- Percent of Owner Occupied Housing Units
- Percent of Population Living in Units with 1.01 Person or More Per Room
- Percent of Population Living in Units Lacking Some Plumbing
- State Aid Ratio
- Prior Achievement

The actual equations, their parameters, scatter plots, and graphs are available in the technical manual referred to above.

While the PIE equations were developed using a stepwise technique, the process variables were studied using the Full and Restricted Models approach as described by Bottenberg and Ward[4]. The Full Model for a given criterion includes the socioeconomic variables of the original equation plus the process variable of interest. The Restricted Model is the original equation without the process variable. The percents of variance accounted for on the criterion by the Full and Restricted Models are compared. The difference is the unique portion of variance which can be attributed to the process variable. An F test can be applied to test the statistical significance of this unique contribution as follows:

$$F = \frac{(R_1^2 - R_2^2)/df_1}{(1 - R_1^2)/df_2}$$

where:

R_1^2 = the squared multiple correlation of the Full Model (i.e., the percent of variance on the criterion accounted for by all the variables).

R_2^2 = the squared multiple correlation of the Restricted Model (i.e., the percent of variance on the criterion accounted for by the original PIE equation).

df_1 = the number of linearly independent variables in the Full Model less the number in the Restricted Model.

df_2 = the number of cases or observations less the number of linearly independent variables in the Full Model.

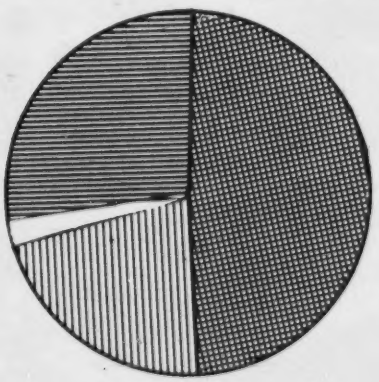
The simple correlation of a process variable to each control variable must also be considered due to the problem of colinearity.

The methodology described above is most conservative. The unique contribution alone is tested. Any commonality of process variables with socioeconomic factors is attributed to the control variables. The Coleman Report[5] has received extensive criticism for using a stepwise approach that does not take into effect the colinearity between background and school process. Figure 1, representing real data on teacher certification, demonstrates the problem. In 1971 the unique contribution of a teacher certification variable to the criterion was statistically significant; however, the magnitude of the contribution was quite small. When one visualizes the statistical relationships in graphic form, as suggested by Mayeske et al.[6] in a reanalysis of the Coleman data, it is apparent that the first-order correlation (represented in Figure 1 by unique plus shared variance) of percent uncertified teachers with performance is of a much greater magnitude than the unique contribution alone. This cannot be noted in the regression technique described, since most of this variance is common to or shared with the background variables.

Figure 1

1970 Percent Uncertified Teachers and
1971 Third-Grade Mathematics With
Socioeconomic Factors Controlled

Total Variance on Third-Grade Mathematics



Unaccounted for variance

Variance uniquely accounted for by socioeconomic factors

Variance uniquely accounted for by percent uncertified teachers

Common or shared variance

The method described above was applied to fifteen school district variables. Results were reported in several Bureau papers.[7,8,9,10,11,12,13,14] The major findings are described below.

Attendance Rate

A significant positive relationship was found between district attendance rate and achievement; however, when socioeconomic factors were controlled, the relationship disappeared. This suggests that the relationship is primarily a result of socioeconomic differences between districts.

Mobility

District rate of student mobility showed a significant negative correlation with achievement in a study of 80 districts. The fact that the relationship disappeared when socioeconomic factors were considered indicates that, in districts of similar socioeconomic makeup, mobility is unrelated to achievement.

Enrollment

Enrollment was negatively correlated with achievement, but when district population was entered as a control variable, enrollment made no unique contribution. This suggests that enrollment is an indicator of community type rather than a factor that can be manipulated to improve achievement.

Teacher Characteristics

No relationship was found between median age of teachers in a district and student achievement as measured by mean district PEP scores. The same is true for median years of teacher experience. Analyses of the relationships of percent of married teachers and percent of uncertified teachers resulted in contradictory findings in different years.

A negative relationship was found between percent of male teachers in a district and 16 of 28 PEP tests and subtests.

Finally, the positive relationship between achievement and percent of teachers with graduate training remained after accounting for socioeconomic factors.

Care should be used in interpreting these results. A school superintendent considering hiring a teacher with 20 years of successful teaching experience should not automatically reject the applicant because median experience does not correlate with achievement. The results are more likely to be useful in developing policies regarding the hiring of teachers and encouraging graduate training than in making decisions about individual teachers.

Similarly, the fact that low achieving schools tend to have a higher percentage of male teachers even when socioeconomic status is controlled for does not imply that male teachers in any way cause low achievement. It does suggest that the nature of that relationship should be more thoroughly examined by educational researchers.

Finally, the strong relationship between percent of teachers with graduate credits and achievement tends to confirm the conventional wisdom regarding teacher training. On the basis of these findings, it is not unreasonable to suggest that districts continue providing incentives for teachers to take graduate courses.

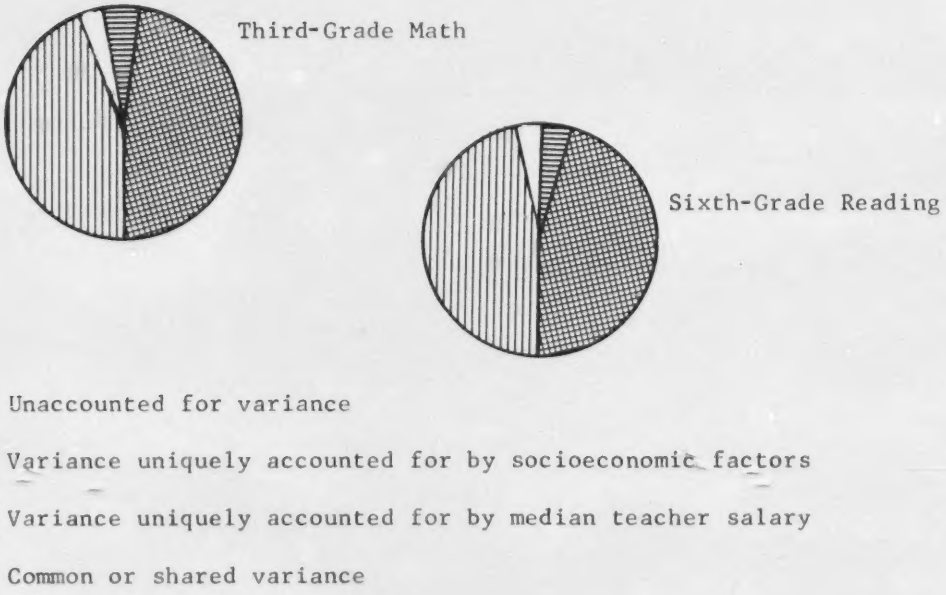
Teacher Salaries

On all reading and mathematics tests and subtests, median teacher salary showed a unique relationship beyond that accounted for by socioeconomic variables, suggesting that salaries are not merely a reflection of a district's ability to pay. Figure 2 illustrates two of these relationships. Principals' salaries also appear to be related to performance.

Figure 2

Relationship of Median District Teacher Salary
To Reading and Mathematics Achievement
With Socioeconomic Factors Controlled

Total Variance on 1971 PEP


Third-Grade Math
Sixth-Grade Reading
Unaccounted for variance
Variance uniquely accounted for by socioeconomic factors
Variance uniquely accounted for by median teacher salary
Common or shared variance

Conclusion

The findings summarized above indicate the inexactness of our knowledge of school processes. Additional results of the regression studies are not reported here because of their ambiguity. Even fairly straightforward and consistentresults should be interpreted cautiously for several reasons:

1. The findings show relationships between variables but not what causes what.

2. The findings are based on one set of school districts (in most cases, all the districts in New York State) at one or more times in the past. The same results may not hold for other school districts or for other time periods.

3. The findings show relationships which have been found for the school districts as a group, but the relationships may not hold true for any individual district.

Nevertheless, the findings provide a starting point for considering changes which are intended to improve the performance of pupils in two areas of the curriculum. The findings can be used to supplement knowledge obtained from other sources. For example, a school superintendent who is trying to decide between two candidates for a position may want to consider the findings on graduate training and experience, but also he would certainly use his knowledge of other qualities needed in the position in making his decision.

It is hoped also that the findings may serve as a starting point for additional research. Research is needed to analyze more thoroughly school processes, to obtain more satisfactory indicators of student performance, and to establish causality between school processes and student performance.

PART III

THE OUTLIER STUDY

PART III

THE OUTLIER STUDY

The research reported in Part II used the school district as the unit of study. This part of the report describes the strategy used and the results obtained in a study of school buildings.

A Strategy for Studying School Processes

One approach to the study of school processes is to compare "good" schools with "poor" schools on a number of school factors. But identifying good and poor schools has been a problem.

School quality is sometimes defined in terms of the performance of the students in a school. However, research by a number of investigators has indicated the extent to which students' performance is related to factors over which the school has little or no control. Perhaps the most well known of this research is the Coleman Report.[5] Therefore, examining the performance of students without considering those factors not controlled by the schools is likely to produce inaccurate information about the role the school played in determining students' performance.

The major thrust of the PIE project has been to attempt to evaluate districts after controlling for differences in these influential community factors. It seemed reasonable, then, that "good" and "poor" schools might be identified by the degree to which they exceed or fail to attain the level of achievement which community factors indicate is reasonable to expect.

The strategy decided upon to test this hypothesis included the following steps:

1. Compute expected levels of achievement for schools in the state, using socioeconomic variables and other factors not controlled by the schools.

2. Identify those schools with actual average achievement scores which vary most from their expected achievement, both positively, (high-performing schools) and negatively (low-performing schools). These schools have been dubbed "outliers."

3. Compare high-performing schools with low-performing schools on a number of variables which describe school processes.

Determining Performance of Schools

A contract was negotiated with two faculty members of the State University of New York at Buffalo, through the Faculty of Educational Studies, to identify schools with exceptionally high and exceptionally low levels of performance and to investigate characteristics associated with performance. Their findings were included in two reports to the State Education Department.[15,16]

The first step in determining performance of schools was to develop criterion measures. A principal components factor analysis was made of school mean subtest scores for third- and sixth-grade reading and mathematics scores on the Pupil Evaluation Program (PEP) tests for the four school years 1969-70 through 1972-73. The analysis revealed a general factor which accounted for 80 percent of the total variance in school means on all subtests in all years. All subtests had correlations with the general factor greater than .80. This indicated that no large amount of variance among school means on any test in any year would be lost by retaining only one factor.

Rather than develop an exact measure of the general factor, an effort was made to develop a simple indicator which was clearly identified with the underlying influences. This was accomplished by using the average of the reading and mathematics test scores for third and sixth grades as a general factor score. Only PEP scores for 1972-73 were used to avoid unnecessarily reducing the number of schools that could be studied. This procedure produced criterion measures for 2,624 schools. Included were public and private schools which contained both a third and a sixth grade.

The strategy used to obtain an expected level of achievement on the criterion measure was to control for variables which could be considered to be indicators of student background and were relatively immune to change as a result of the schools' efforts. A number of such background variables were identified in the data routinely collected by the State Education Department through its Basic Educational Data System (BEDS). Among them were:

1. Whether the school is public or private.

2. The percent of Black students in the student body.

3. The percent of Spanish-American students.

4. The percent of students from families primarily supported by public assistance.

5. Geographical location of the school: a) New York City; b) urban area other than New York City; c) suburb of a major city; or d) primarily rural.

Multiple regression analysis was used to determine how well background variables predicted the criterion measure. Each of the variables listed above contributed significantly to the prediction with the result that a regression equation was formed which accounted for 69.6 percent of the variance of the criterion measure.

Using the equation, a predicted criterion score was calculated for each of the 2,624 schools with both third- and sixth-grade PEP scores. The predicted criterion score was subtracted from the actual criterion score for each school to obtain a residual score. Thus, a high positive residual score indicated that a school was achieving better than expected from the background variables. A high negative score meant the opposite. Since the criterion measure was developed from reading and mathematics test scores for third and sixth grades, the residual was interpreted as a generalized measure of performance in the basic academic skills at the elementary school level.

The difference between the actual and predicted scores was used to identify outlying schools. Schools which had a residual score in excess of +5 were classified as positive deviates. Those having a residual greater than -5 were classified as negative deviates. Schools which had a residual score between .5 and -.5 were classified as non-deviates. The standard deviation of the residual scores was 3.161.

The three groups of schools were subjected to a one-way analysis of variance on 177 variables relating to pupil background, school location, professional personnel, program, and achievement. Parallel analyses were made for private and public schools combined and for public schools alone. Personnel data were not available for private schools.

Findings

One hundred forty-eight schools were identified as positive deviates. Of these, 43 (29%) were private schools; 105 were public. In the negative deviate group, there were 145 schools. Fifty (34%) were private and 95 were public. In the non-deviate group, 104 (32%) were private and 219 were public.

Table 1 reports the means and standard deviations for the three groups on the criterion variable and on the control variables (the independent variables in the equation predicting achievement). As would be expected, there is little variation among the three groups within each analysis on the control variables. The total group averages about 12% Black enrollment with approximately 1/6 of the schools exceeding 30%. Spanish American enrollment averages over 6% with approximately 1/6 exceeding 20%. The public schools average more than 12% of their pupils on welfare; the percentage is lower for

Table 1

Means and Standard Deviations of Criterion and Control Variables
for Six Groups of Deviate and Non-deviate Schools

Control and Criterion Variables	Public and Private Schools						Public Schools Only					
	Positive Deviate		Non-Deviate		Negative Deviate		Positive Deviate		Non-Deviate		Negative Deviate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Public-private[1]	1.29	.45	1.32	.47	1.35	.48	1.00	.00	1.00	.00	1.00	.00
Percent Black	10.79	18.00	9.14	19.83	12.25	22.15	12.83	19.62	12.77	23.08	12.55	22.32
Percent Spanish American	4.65	12.93	5.58	14.13	4.04	10.13	5.95	14.85	7.67	16.52	5.52	12.08
Percent families on welfare[2]	2.69	2.25	2.86	2.14	2.72	1.39	3.08	2.48	3.38	2.37	3.12	1.37
New York City[3]	-.05	.61	-.12	.60	-.03	.60	.03	.65	-.08	.68	.05	.67
Other city[4]	.14	.73	-.07	.65	.13	.71	.12	.71	-.13	.63	.07	.68
Suburb of large city[5]	.05	.67	.12	.81	.07	.68	.07	.70	.20	.81	.11	.70
Criterion of Achievement	41.24	4.87	34.88	4.99	28.14	3.72	40.00	5.02	33.35	5.31	27.21	3.65

[1] 1 = public; 2 = private

[2] 1 = none; 2 = 1-10%;3 = 11-20%; 4 = 21-30% etc.

[3] 1 = New York City; 0 = center city or suburb; -1 = other location

[4] 1 = other center city; 0 = New York City or suburb; -1 = other location

[5] 1 = suburb of large city; 0 = New York City or other center city; -1 = other location

private schools. Private schools also tend to enroll a smaller percentage of students from minority groups. On the achievement criterion, positive deviates average approximately one standard deviation above the total group mean; negative deviates average approximately one standard deviation below. The average for non-deviates falls near the total group mean. Private schools on the average achieve slightly higher than public schools.

Table 2 lists those school and program variables with a difference among group means which is significant at a .05 level or above. Also reported are group means and levels of significance of differences between individual group means and the respective means of the other two groups combined.

Both positive and negative deviate schools average about 100 students less than the non-deviates. Regardless of classification, private schools, averaging 220 pupils, are much smaller than public schools, averaging 584 pupils. Generally, as the analysis shifts from the total group of schools to public schools only, the direction of difference in school and program characteristics remains the same but in some instances the magnitude of the difference drops. It would appear that private schools more than public schools tend to organize programs around the academically talented, multi-age grouping, and non-graded classes.

Positive deviates are more likely to have programs for the academically talented, an open classroom arrangement, and multi-age grouping. They have significantly fewer compensatory education programs and guidance counselors. The number of students per classroom is likely to be lower for both groups of deviates. Traditional classrooms are likely to be found in most schools. However, the proportion of traditional classrooms is lowest in the negative deviate schools. The non-deviate schools are generally less likely to have instituted what might be termed "innovative" programs than either of the deviate groups.

The differences in achievement statistics are consistent for the four years studied, 1969-70 to 1972-73. Findings with reference to achievement are reported in Table 3 for 1971-72 only. For all years, all grades, and all subjects, achievement means are high for the positive deviate schools and low for the negative deviate schools. The standard deviations are smaller for positive deviates than for negative deviates for all years and for all subjects except sixth grade mathematics where the reverse is true for all years. Most differences on the achievement means are significant at the .001 level.

Table 2

Means and Levels of Significance of School and Program Variables
Which Differentiate Among the Groups with a Chance Probability of Less Than .05

School and Program Variables	Public and Private Schools[1]			Public Schools Only[1]		
	Positive Deviates	Non-Deviates	Negative Deviates	Positive Deviates	Non-Deviates	Negative Deviates
Total enrollment	426	517***	402**	525*	637***	526*
Academically talented	.257***	.149	.083**	.210	.187	.084*
Compensatory program	.162*	.245	.290	.171**	.311	.305
Students/total rooms ratio	16.80**	18.87***	16.49***	18.60	19.21*	17.87*
Students/regular classroom ratio[2]	26.88**	28.74*	27.71	27.87	28.89	28.53
% Total rooms regular classrooms	63.05	65.52***	59.92***	66.89	66.79*	62.56***
Traditional classroom arrangement	.939	.957*	.890**	.943	.973*	.905*
Open classrooms[2]	.243**	.115**	.172	.248	.146	.200
Multi-age grouping[2]	.324*	.204**	.290	.267	.219	.274
Non-graded classes[2]	.223	.164**	.317***	.210	.160	.263
Guidance counselors	.270*	.341	.461**	.276***	.411	.589***

[1]Individual group means that differ from the mean of all other groups combined are indicated by asterisks. The number of asterisks indicates the probability level: three (***) = .001; two (**) = .01; one (*) = .05.

[2]Not significant for public schools only.

Table 3

Means and Levels of Significance of Achievement-Related Variables
for 1971-72 Which Differentiate Among Groups With a Chance Probability of Less Than .05

Achievement-Related Variables	Public and Private Schools[1]			Public Schools Only[1]		
	Positive Deviates	Non-Deviates	Negative Deviates	Positive Deviates	Non-Deviates	Negative Deviates
Means						
Grade 3 - Reading	36.78***	32.92	27.28***	35.73***	31.50	25.79***
Grade 3 - Mathematics	36.96***	32.25	26.87***	36.24***	30.97	25.75***
Grade 6 - Reading	46.00***	41.29	36.30***	45.06***	39.30	34.59***
Grade 6 - Mathematics	37.42***	32.35	27.76***	37.00***	30.96	26.53***
Standard Deviations						
Grade 3 - Reading	9.78***	10.36	10.93***	10.28***	10.84	11.35***
Grade 6 - Reading	11.31***	11.82	12.28**	11.90***	12.47	12.77*
Grade 6 - Mathematics	11.46***	10.69	9.98***	11.83***	10.98	10.11***

[1]Individual group means that differ from the mean of all other groups combined are indicated by asterisks. The number of asterisks indicates the probability level: three (***) = .001; two (**) = .01; one (*) = .05.

Data on the characteristics of teaching staffs are available only for public schools. The findings are reported in Table 4. Positive deviate schools tend to have a smaller percentage of male teachers than do the other two groups (16%, 20% and 22% respectively). They also tend to have more highly trained staffs. Non-deviates have the fewest uncertified teachers, 3.3%, while negative deviates have the most, 5.6%. Positive deviates average 4.1% uncertified teachers but they have the largest standard deviation, 10.7%. A larger percentage of teachers in positive deviate schools are on tenure than for the other two groups (69%, 65% and 59% respectively). Teachers in positive deviate schools have more experience in the district and in total years teaching. The means in average total experience for the three groups are 12, 11, and 10 years respectively. Teachers in positive deviate schools earn the highest salaries. Their counterparts in the negative deviate schools earn the least.

Conclusions

The three groups of schools are traditionally oriented; however, deviate schools, both positive and negative, tend to be more innovative than non-deviate schools. Those who still believe that schools can make a difference in children's learning can gather some comfort from the analysis of teacher characteristics. Teachers in the positively deviate schools as a group epitomize what conventional wisdom claims to be characteristic of good teaching staffs. They are better trained and more experienced. They are paid higher salaries and are more likely to be on tenure. This still does not inform us as to what well trained, experienced and highly paid teachers <u>do</u> to bring about unexpectedly high pupil achievement. Such knowledge is essential if we are to prepare and organize teachers and the teaching processes to obtain optimum results.

Table 4

Means and Levels of Significance of Variables Describing Characteristics of Teaching Personnel Which Differentiate Among the Groups With a Chance Probability of Less than .05 for Public Schools Only

Teacher Characteristic Variables	Positive Deviates[1]	Non-Deviates[1]	Negative Deviates[1]
Number of teachers	24.76**	30.64***	26.98
Percent male teachers	16.24***	19.77	22.28**
Percent uncertified teachers	4.06	3.27*	5.57*
Percent Bachelors degree only	42.01**	48.89	51.45*
Percent BA + 30 or Masters degree	44.26**	39.45	37.51*
Percent MA + 30 or Doctorate	11.99*	9.04	9.01
Percent tenure	69.28**	64.53	59.49**
Percent full time	97.44	97.11	95.67
Mean degree status	6.07***	5.89	5.86
Mean years experience in district	9.23**	8.54	7.91*
Mean total years experience	11.90***	10.84	9.86***
Mean salary	$12,740.00***	$12,061.00	$11,405.00***
Mean age	39.01***	37.24	36.18**

[1]Individual group means that differ from the mean of all other groups combined are indicated by asterisks. The number of asterisks indicates the probability level: three (***) = .001; two (**) = .01; one (*) = .05.

PART IV

THE OBSERVATIONAL STUDY

An Ecological Study of Teachers and Administrators in Fourteen Schools

PART IV

THE OBSERVATIONAL STUDY

Background

The Performance Indicators in Education (PIE) project demonstrated that all schools with low socioeconomic backgrounds do not do poorly. An analysis of 1970 Pupil Evaluation Program (PEP) reading and mathematics scores by the Bureau of School Programs Evaluation revealed a wide range of test scores among New York City schools, even those drawing largely from low socioeconomic populations.

An investigation was initiated to determine the extent to which school environment, school program, and administrative conditions might be related to test results. For the study, a group of five schools was identified in which reading scores were consistently high and five schools in which scores were consistently low. Also, two additional schools were identified for study. In one, test scores had increased over a two-year period. In the other, scores had declined.

The twelve schools were then used as the targets for structured observations, carried out under contract with Assessment Associates of Cortland, New York. On the basis of their observations, members of evaluation teams classified eight out of ten of the high and low schools correctly. In seven out of ten cases, all evaluators judged correctly. (Generally, three evaluators visited each school.) While specific differences between high and low scores were not completely consistent, seven factors emerged as likely to be more true of high-achieving schools than of low-achieving schools. Briefly, these characteristics were:

1. Teachers manifested better rapport with students.

2. Teachers exercised more effective control of pupils.

3. Teachers engaged in more extensive preparation of lessons.

4. Reading instruction was at a more appropriate level for the needs of pupils.

5. Teachers provided for more extensive regrouping within the reading period.

6. Teachers provided for more extensive use of material in the reading program.

7. The sources of leadership in instruction in reading were more forceful and positive.[17]

A second study was undertaken to:

1. Verify the findings of the earlier study.

2. Produce more refined instruments for studying school processes.

3. Expand the study to include non-urban school districts.

4. Eliminate several design problems of the earlier study.

The study was carried out under the direction of a faculty member of the State University of New York at Albany. A description of the study and findings was included in a report to the State Education Department.[18] The major elements of that report are presented below.

Instrument Development

As a first step, the original forms used in the initial study were examined. It was agreed that various items of demographic data such as the socioeconomic level of pupils would not be stressed in the instruments to be developed. Rather, more attention would be given to aspects of teacher behavior and school environment that are under the control of the school. If any of the aspects noted were to be related in a meaningful way to differences in pupil learning, such aspects would be something that educators might do something about. From the original forms, a list was made of items for further consideration. The importance of and the psychometric problems connected with each item were considered, as well as ways and means of gathering relevant information. Specific and general scales were constructed, edited, and used in field tryouts in two schools. In each case, four observers visited the school. Pairs of observers made twenty-minute observations in six different classrooms so that each observer was paired with every other observer. Also a pair of observers interviewed the reading teacher and the other two observers interviewed the school principal. After the field trials, data were inspected for reliability and other considerations, and the forms were revised.

Eight different instruments were developed. Four of the instruments were designed to be used in each of the classrooms visited. One observation instrument called for a general assessment of the school. Also developed were two interview schedules and a questionnaire to be filled out by the teachers observed. The names of the instruments and a brief description of each follow.

1. General Classroom Observation - This form was designed for rating the degree of existence of various aspects of classroom emphasis, teacher behavior, pupil behavior, and facilities. It includes 16 five-point rating scales.

2. Teacher Questionnaire - This multiple choice form was designed to elicit teachers' perceptions of their students, the program, and administrative support.

3. Principal Interview - This instrument contains a set of five-point rating scales for eliciting principals' opinions regarding teacher effectiveness, adequacy of facilities, and degree of their rapport with teachers, parents, students, and school board.

4. Observation of a Reading Group - The form is a modification of a system developed by Educational Testing Service.[20, 21] Sixteen categories describing the reading program are scored on a Likert scale arranged from "little" to "much."

5. Reading Teacher Interview - The interview guide was adapted from an "Observer Guide--Reading," published by the Bureau of Reading, New York State Education Department. Thirteen categories to be evaluated from "low" to "high" on a five-point scale were selected to describe the degree to which reading practices were seen as ideal.

6. Teacher Reinforcement Scale - Likert scales from "low" to "high" were used to described the frequency and strength of both negative and positive reinforcement provided by the teacher.

7. Characteristics of Open Education - This form is a shortened version of a scale developed by Walberg and Thomas[22] to operationalize the definition of open classroom. It contains 18 items using five-point Likert scales to measure the degree of openness.

8. General School Observation - This form, using a Likert scale, calls for an overall assessment of morale, level of expectancy, and general school appearance.

Procedures for School Visits

After instruments had been revised and constructs redefined, fourteen school buildings were selected on the basis of the Outliers Study described in Part III of this paper. Seven of the buildings were high outliers (positive deviates) and seven were low outliers (negative deviates). At this stage, the observers were unaware of whether the schools they visited were "high" or "low."

For each school visit, the principal was asked to arrange access to nine elementary school classrooms between kindergarten and sixth grade. (In a few of the smaller schools, nine classrooms were not available.) Each observer was assigned to four of these classrooms, and the pair of observers was together in one classroom during the day. In addition, interviews were held with the school principal and with a reading teacher in the school. The original design called for each observer to be paired with every other observer, but because of scheduling problems, this plan was not completely carried out.

Analysis of Data: Overview

In the following seven sections of this report, data are presented on seven of the eight instruments developed for the study. For each instrument, mean scores are presented for (1) first- to third-grade classes in the seven high-performing schools; (2) first- to third-grade classes in the low-performing schools; (3) fourth- to sixth-grade classes in the high-performing schools; and (4) fourth- to sixth-grade classes in the low performing schools. Tests of significance are not reported for each pair of means, nor are standard deviations provided. However, almost all data were obtained on five-point Likert scales. Most standard deviations were very close to 1.00. In general, group sizes were similar. As a rule of thumb, differences between means of approximately .50 points can be considered to be significant at the .05 level.

In addition to mean scores, certain of the correlational relationships are displayed and commented upon. With the great number of variables in the study, not all of the correlations are presented.

Mann-Whitney U tests were used on selected segments of several devices in order to ascertain whether the devices significantly discriminated, on a rank order basis, between high and low schools.

Finally, a root mean square procedure was used to test the extent of agreement between raters in this study.

General Classroom Observations

Data from the General Classroom Observation form reported in Table 5 show clear differences between high-performing and low-performing schools in grades 1-3 but relatively little difference in grades 4-6.

For early elementary grades, items which did not differentiate between the groups of schools were program emphasis on social development, rigidity of student behavior, and three items on facilities. High-performing schools were significantly higher on all items except "effort to maintain control," and "rigidity of student behavior." For these ratings, high-performing schools were significantly lower. Thus, in grades 1-3 classes in high-performing schools, teachers apparently exerted less effort to maintain control and had less rigid student behavior, but at the same time were rated significantly higher in effectiveness of control. Also teachers in high-performing schools were rated as warmer, more supportive, more responsive to students, and showing more emphasis on cognitive development. Pupils in their classes appeared more enthusiastic about school and better able to sustain attention.

In grades 4-6, however, only the differences in teacher efforts to maintain control and effectiveness of control were significant. The pattern was the same as in the lower grades, with teachers in high reading schools making less effort to maintain control but being rated more effective in control.

Table 5

Means for High- and Low-Performing Schools
on General Classroom Observation Form

Item Number	Grades 1-3		Grades 4-6		Kindergarten	
	High Performing	Low Performing	High Performing	Low Performing	High Performing	Low Performing
Emphasis on:						
1.Cognitive Development	2.9	2.2	2.9	2.5	2.8	2.7
2.Language Development	3.2	2.5	2.7	2.7	2.8	2.2
3.Social Development	2.3	2.2	2.1	2.1	3.0	3.2
Teacher Behavior:						
4.Effort to Maintain Control	1.7	2.5	1.5	2.0	1.8	2.0
5. Effectiveness of Control	3.9	3.3	3.7	3.3	3.5	3.8
6.Warmth of Pupil-Teacher Interaction	3.2	2.7	2.7	2.9	3.1	4.2
7.Amount of Interaction with individual pupils	3.2	2.7	3.0	2.6	3.3	3.3
8.Amount of supportive or risk reducing activity	2.6	2.2	2.0	2.1	2.9	3.5
9. Responsiveness of pupil's ideas	2.9	2.1	2.4	2.3	2.1	3.0
10.Use of pupil responses to guide teacher's strategy	2.7	1.9	2.3	2.3	2.2	2.5
Pupil Behavior						
11.Apparent enthusiasm for school	3.4	2.9	2.7	2.9	3.6	3.8
12.Rigidity of pupil behavior	3.1	3.8	3.6	3.4	2.6	3.7
13.Ability to sustain attention	3.6	2.8	3.4	3.0	3.3	3.7
Facilities						
14.Attractiveness of room displays	3.1	2.8	2.3	2.6	3.3	3.3
15.Adequacy of space	3.2	3.4	3.1	3.1	3.8	4.2
16.Use of space	3.1	2.8	2.8	2.4	3.4	4.0

Teacher Questionnaire

Teacher questionnaires were left in each of the schools visited, and a request was made to return the questionnaires when completed to the Project Director. Forty-eight responses from high-performing schools and 51 responses from low-performing schools were available for analysis. The first four items of the questionnaire concerned teachers' expectations for children in their classes. These data are reported in Table 6.

Table 6

Mean Expectation for Pupils Expressed
by Teachers in High- and Low-Performing Schools
(Nine-point scale)

Expectation for Pupil to	High-Performing Schools N=48	Low-Performing Schools N=51
Graduate from high school	8.85	7.36
Go to college	6.09	2.73
Become fluent reader	7.17	5.66
Become a good adult citizen	8.72	7.89

As can be seen, teachers in high-performing schools had higher expectations for children in all four of the areas questioned. It is interesting to note in each group that teachers expected more children to graduate from high school than they expected to become fluent readers.

A similar pattern is seen in Table 7 which reports the results of the questionnaire in which teachers were asked their present perceptions of their pupils. Teachers in high-performing schools saw children in their classes as more intelligent, better behaved, more pleasant to teach, and having more concerned parents than did teachers in low-performing schools.

Also derived from the teacher questionnaire were teachers expectations for various kinds of support within the school. These data are provided in Table 8.

Table 7

Mean Perception of Pupils Expressed by
Teachers in High- and Low-Performing Schools

Teacher's Perceptions of Pupils	High-Performing Schools N=48	Low-Performing Schools N=51
Intelligence	3.40	2.75
Behavior	3.27	2.86
Pleasant to teach	3.94	3.45
Concern of parents	4.02	3.09

Table 8

Mean Expectation for Support Expressed by
Teachers in High- and Low-Performing Schools

Expectation for Support in	High-Performing Schools N=48	Low-Performing Schools N=51
Getting needed instructional material	3.75	3.49
Dealing with a behavior problem	3.13	3.33
Developing my own teaching skill	2.66	2.75
Dealing with specific learning problems	3.19	3.26

As can be seen, responses of teachers in high-performing and low-performing schools were very similar. It appears from these data that, difference in administrative support as perceived by teachers is not a critical factor.

Principal Interview

The principal in each school was interviewed through the use of a relatively unstructured procedure. The interviewer then attempted to score a number of responses. One set of items related to the principal's evaluation of rapport and competency of various staff members. These data are reported in Table 9. As can be seen, principals in high-performing schools provided consistently higher ratings than did principals in low-performing schools. All average ratings fell above the midpoint of the Likert scale, indicating that, on the average, principals gave favorable ratings.

Table 9

Principal Ratings of Personnel
as Ascertained from Interviews

Principal Ratings of	High-Performing Schools	Low-Performing Schools
Teacher rapport	4.50	4.14
Teacher competency	4.57	3.86
Reading teacher competency	4.43	3.57
Librarian competency	4.60	4.29

Principals were also asked about their own rapport with various groups with whom they deal. These data are summarized in Table 10. Principals in low-performing schools were rated from the interview as higher in rapport with the school board, but lower in rapport with teachers, parents, and children. Especially noticeable was the difference of the two groups in ratings for rapport with children. However, because the number of principals in each group was only seven, conclusions must be tentative.

Other data, available from the Principal Interview form, related to differences in program organization, structure of the reading program, and the like. In general, these variables did not differentiate high- and low-performing schools; therefore, they are not reported here.

Table 10

Principal Ratings of Rapport as Ascertained from Interviews

Rapport with	High-Performing Schools	Low-Performing Schools
School board	3.83	4.33
Teachers	4.57	4.00
Parents	4.43	4.14
Students	4.43	3.43

Observation of Reading

The Observation of Reading form was designed to be used when the teacher was engaged in direct reading instruction. When arrangements for school visits were made, a desire was expressed to observe some reading classes, if possible, but also to see other activities. Since observers were in classes most of the school day, and reading was normally scheduled in the morning, reading classes were often not available. In Table 11 the mean ratings are presented for activities carried on in reading classes in high- and low-performing schools. As can be seen, relatively few reading classes were observed in grades four to six.

In grades one to three, more total activity seemed to occur in the high-performing schools. On nine of the eleven activities more emphasis was rated for high-performing schools than in low-performing schools. The two areas in which higher means were recorded for low-performing schools were for reading orally and for management instructions. The greatest difference between high- and low-performing schools was in the relatively large amount of silent reading going on in high-performing schools. Thus, in classes in high-performing schools, children were observed more often reading silently; in low-performing schools more children were observed reading out loud. Although the number of classrooms was very small, this relationship between silent and oral reading extended to grades four to six.

Table 11

Mean Ratings of Activities Carried on in Reading Classes in High- and Low-Performing Schools

Activity	Grades 1-3		Grades 4-6	
	High-Performing Schools N=34	Low-Performing Schools N=38	High-Performing Schools N=23	Low-Performing Schools N=25
Comprehension	3.22 (18)	2.45 (22)	3.50 (8)	3.33 (3)
Pronunciation and Word Recognition	3.32 (19)	2.92 (27)	3.50 (6)	4.50 (2)
Language Structure	2.69 (16)	2.39 (23)	3.17 (6)	2.75 (4)
Reading Silently	3.75 (12)	2.47 (15)	3.20 (5)	2.50 (2)
Reading Orally	3.40 (15)	3.68 (22)	2.50 (4)	4.50 (2)
Spelling	2.92 (12)	2.23 (17)	2.40 (5)	5.00 (1)
Writing	2.25 (4)	1.00 (6)	1.00 (1)	-- (0)
Copying	2.40 (5)	1.63 (8)	2.00 (3)	3.50 (2)
Listening Instructions	2.18 (11)	1.00 (7)	2.60 (5)	3.00 (2)
Non-reading Instructions	1.75 (4)	1.55 (11)	1.00 (2)	1.50 (2)
Management Instructions	1.69 (13)	2.25 (27)	1.60 (5)	1.50 (2)

*Numbers in parentheses are the actual number of classrooms observed for that activity under each category.

Also examined in the Observation of Reading Form were the behaviors of children who were not in reading groups. These data are reported in Table 12. Again, it should be noted that the number of classes involved in reading was small, so the differences seen are not necessarily statistically significant.

Table 12

Means of Ratings of Activities of
Non-Reading Groups During Reading Period

Activity	Grades 1-3		Grades 4-6	
	High-Performing Schools	Low-Performing Schools	High-Performing Schools	Low-Performing Schools
Reading Silently	3.00	3.30	3.27	2.00
Writing	2.56	2.33	2.67	2.50
Copying	2.92	2.25	2.86	3.50
Non-reading instruction	2.77	2.84	3.25	4.50
Play	1.75	2.10	1.00	1.50
Social Interaction	2.11	2.18	2.00	1.60

Reading Teacher Interview

An interview was held with a reading teacher in each of the schools visited. The reading teacher was asked to evaluate a number of aspects of the classroom reading program. These data are reported in Table 13. It should be noted that the averages in the table are based on the responses of the seven reading teachers in the high-performing schools and the seven in the low-performing schools. Although the small sample size did not permit statistical analysis, a definite trend can still be seen since higher means were recorded for the high-performing schools in nine of the ten contrasts. Only on item six, which concerns the degree to which reading material matches the background of the child, were low-performing schools given a higher mean.

Also of interest were the absolute scores obtained from the five-point Likert scale. In general, absolute scores were higher than the scores assigned through direct classroom observation. It is not surprising that reading teachers should have generally seen the program in their school in a favorable light. Nevertheless, especially in the low-performing schools, most ratings fell between 3 and 3.5, suggesting that reading teachers saw room for improvement.

Table 13

Mean Responses of Reading Teachers in High- and Low-Performing Schools

Aspects of Reading Programs	High N=7	Low N=7
1. Variety of material	3.67	3.25
2. Appropriateness of use	4.17	3.13
3. Effort to extend reading	3.67	3.00
4. Efforts to ask questions	3.16	3.00
5. Efforts to give purpose	4.20	3.00
6. Relationship of material to childrens' backgrounds	3.00	3.50
7. Use of informal diagnosis	4.33	3.29
8. Flexibility of grouping	3.83	3.38
9. Effective use of test data	3.50	3.38
10. Availability of interesting books	4.33	2.88

Teacher Reinforcement Scale

The Teacher Reinforcement Scale offered a record of the frequency and strength of positive and negative reinforcement. Ratings were made for reinforcement specifically related to instruction and for general support. Data are reported in Table 14. As can be seen, teachers in grades one to three in high-performing schools provided more positive reinforcement under all conditions, and teachers in low-performing schools provided more negative reinforcement. However, in grades 4-6 teachers in low-performing schools provided more reinforcement, both positive and negative, than did teachers in high-performing schools.

In the instructional-specific category, teachers in all groups were recorded on the average as providing more positive reinforcement than negative reinforcement. Under the category of general support, however, teachers in high-performing schools had a higher frequency of positive than negative reinforcement, while teachers in low-performing schools showed the reverse pattern. Only two mean ratings reached the midpoint of the five-point scale. This indicates that, in general, the observers did not see a great amount of reinforcement.

Table 14

Mean Scores for Teacher Reinforcement in Classes in High- and Low-Performing Schools

Type of Reinforcement		Grades 1-3		Grades 4-6	
		High-Performing Schools	Low-Performing Schools	High-Performing Schools	Low-Performing Schools
Instructional-specific					
Frequency	Positive	3.08	2.58	2.27	2.42
	Negative	1.47	1.87	1.68	1.92
Potency	Positive	2.97	2.79	2.31	3.00
	Negative	1.81	2.37	1.86	2.58
General Support					
Frequency	Positive	2.56	1.82	2.09	2.16
	Negative	1.53	2.58	1.52	2.28
Potency	Positive	2.76	2.47	2.33	2.96
	Negative	1.90	2.66	1.76	2.88

Characteristics of Open Education

The form, Some Characteristics of Open Education, is a shortened version of a scale developed by Walberg and Thomas[20] to measure the degree of classroom openness. Since this form was not used in preliminary field work, no data were available to form a basis for revision. Observers reported that some items were not really suitable for observation and also indicated that polar definitions would have been useful.

However, in spite of these problems, a definite trend was seen in scores for high-performing and low-performing schools. Of the 18 comparisons in grades 1-3, 15 were in favor of high-performing schools, as can be seen in Table 15. This trend was not strong in grades 4-6, in which means for high-performing schools were greater in 11 of 18 comparisons. For almost all items of open education characteristics, ratings were below the midpoint of the five-point scale. Thus, one might assume that the general sample of schools employed in this study was relatively traditional, and this assumption was verified by the opinions of the observers.

Table 15

Comparison of Mean Scores of Characteristics of Open Education in High- and Low-Performing Schools

Teacher Characteristics (Abbreviated)	Grades 1-3		Grades 4-6	
	High-Performing Schools	Low-Performing Schools	High-Performing Schools	Low-Performing Schools
1. Time to individuals, not group	2.51	2.45	2.57	2.08
2. Teacher uses child interaction	2.16	1.75	2.06	2.00
3. Teacher encourages independence	1.48	1.71	1.47	1.59
4. Individual child diagnosis	2.30	2.37	2.00	1.82
5. Teacher encourages fantasy	1.67	1.16	1.38	1.16
6. Diverse instructional materials	2.53	1.97	1.75	1.61
7. Materials accessible	2.94	2.57	2.35	2.17
8. Children move freely	2.41	2.13	2.13	1.75
9. Children use other areas	1.11	1.12	1.50	1.24
10. Many diverse activities	2.21	2.02	1.86	1.48
11. Much individual work	1.81	1.63	1.71	1.39
12. Individual evaluation	2.53	2.05	1.79	2.00
13. Individual instruction	2.62	2.07	2.08	2.10
14. Teacher respect for child	2.63	1.97	2.00	2.33
15. Children freely express feeling	2.67	2.29	2.29	2.48
16. Teacher promotes trust	3.04	2.54	2.26	2.57
17. Teacher keeps learning	2.64	2.34	2.22	2.26
18. Teacher is secure	2.09	1.92	1.72	2.00

Correlations of items of the Characteristics of Open Education form were examined, and five items were selected which seemed especially observable in classroom settings. Intercorrelations of these items in grades 1-3 in high- and low-performing schools are presented in Table 16. As can be seen, these items generally were highly correlated, especially for high-performing schools. These intercorrelations suggest that it may be possible to tap a relatively broad characteristic of openness that differentiates classrooms.

Table 16

Correlations of Selected Items Related to Open Education in Grades 1-3

	1	2	3	4	5
1. Time to individuals, not group			.80**	.76**	.81**	.87**
2. Diverse materials, not class sets	.18			.85**	.67**	.75**
3. Materials readily accessible to children	.38*	.57**			.73**	.73**
4. Children move freely without asking permission	.53**	.04	.40**			.84**
5. Many activities going on at same time	.46**	.33*	.49**	.49**	

**significant at .01
*significant at .05

(Classes in high-performing schools are given above the diagonal; classes in low-performing schools are given below the diagonal.)

Discrimination Between High- and Low-Performing Schools

It is desirable to know whether there are clear and reliable differences between schools designated as high and low in performance. In order to assess the consistency of ranking, Mann-Whitney U's were computed on the rank order of high- and low-performing schools on certain variables. For selected items on the Open Education scale, Table 17, there was no significant relationship between achievement category and rank on this scale; high-performing schools were slightly more likely to be rated high on this scale. On the General Classroom Observation scale, Table 18, however, there was a consistent relationship with achievement category ($U = 3.45$), significant at $p < .01$.

High achieving schools generally received higher ranks on this measure, although a few reversals were noted. On selected items from the scale used for observation of reading, Table 19, there was also a significant relationship ($U = 2.81$, $p < .01$) between score on the scale and performance category of the school, with high-performing schools generally having received higher scores.

Table 17

Discrimination of Rank-orders of High- and Low-Performing Schools on Selected Items on Open Education

Score	Rank	Performance Category
15.1	1	High
13.4	2	High
11.8	3	Low
10.9	4	High
10.2	5	Low
10.1	6	Low
10.0	7	High
9.2	8	Low
9.1	9	High
8.7	10	Low
7.7	11	High
7.6	12	High
7.3	13	Low
7.1	14	Low

U = 1.66 N.S.

Table 18

Discrimination of Rank-orders of High- and Low-Performing Schools on Selected Items on General Classroom Observation

Score	Rank	Performance Category
35.30	1	High
33.18	2	High
33.14	3	High
31.30	4	Low
31.20	5	Low
29.56	6	High
29.25	7	High
27.48	8	High
27.81	9	Low
27.30	10	Low
27.23	11	Low
27.10	12	High
25.86	13	Low
24.68	14	Low

U = 3.45 significant at .01

Table 19

Discrimination of Rank-orders of High- and Low-Performing Schools on Selected Items on Observation of Reading[1]

Score	Rank	Performance Category
15.66	1	High
13.00	2	High
13.00	3	High
11.66	4	High
11.60	5	High
11.04	6	Low
11.00	7	High
10.75	8	Low
9.33	9	Low
8.95	10	Low
8.76	11	Low

U = 2.81 significant at .01

[1]Three schools are excluded from this analysis because there were too few occasions in which explicit reading instruction was observed.

Observer Agreement

A number of options presented themselves as procedures which might have been followed to check observer agreement. Separate reliabilities could have been computed for each form, or various techniques of computing observer agreement could have been followed. A simple procedure which seemed adequate at this stage was to obtain root mean squares using the formula:

$$\text{root mean square} = \sqrt{\frac{1}{n} \times \sum_{1}^{n} d^{2}}$$

where n = number of variables and d = difference in ratings between two observers on a variable.

Using this procedure, the possible range of scores is zero to four. A score of zero would indicate perfect agreement on all seventy variables scores on all observational instruments. A score of four would indicate complete disagreement. Table 20 shows the root mean squares obtained for each pair of observers. In general, substantial agreement was found. However, observer one showed considerable disagreement with observers four and six.

Table 20

Root Mean Square Differences Between Observers

Observer	2	3	4	5	6	7
1		1.02	1.81	1.06	1.60	
2			.84	.78	1.20	.97
3				1.25		
4				1.00	1.13	1.13
5						.51
6						.83
7						

It should be noted that the root mean square procedure would show disagreement if observers had different response tendencies, which standard reliability tests would not pick up. In other words, one observer might score very strictly and one quite leniently, but they might agree very well in the profile of their data. The two observers would be credited with high reliability. Under a root square method, however, the absolute difference between their ratings was used which provided a more stringent criterion.

At this stage of instrument development, the degree of relationship found between raters was seen as encouraging, but this should remain an area of future study.

Discussion

In general, the procedures followed in this study seemed to produce meaningful differentiations between schools identified as either high or low in academic performance. Data were generally in harmony with the findings of other studies.

There are several steps that might be taken next:

1. Select the variables that seem to be related to the clearest differences between schools.

2. Attempt to clarify further the behaviors that are being rated and the criteria for rating each of these variables.

3. Consider whether the variables identified by these procedures could be meaningfully divided into sub-parts to be more specifically studied.

4. Try out revised materials on a broader geographic basis and with schools that are demographically more diverse.

5. Experiment with these materials in an in-service and/or pre-service context.

6. Work systematically with a group of teachers to see if teachers can learn to vary selected behaviors, and study the effects of such variations.

Summary of Most Significant Findings

Listed below are seventeen conclusions that emerged from the data. Most of the conclusions focus on differences between the high- and low-performing schools, but a few involve no difference findings. Much could be written on the implications of each of these conclusions, but on another level the conclusions speak for themselves. They should be taken as tentative and subject to further confirmation.

1. Teachers in high-performing schools made less overt effort to maintain class control than those in low-performing schools, had less rigid student behavior, but were more effective in maintaining the level of control they appeared to want.

2. Teachers in high-performing schools were rated as warmer, more responsive, and showing more emphasis on cognitive development in classes that did not involve direct reading instruction as well as in reading classes.

3. Teachers in high-performing schools expected more children to graduate from high school, to go to college, to become good readers, and to become good citizens.

4. Teachers in high-performing schools saw their children as more intelligent, better behaved, more pleasant to teach, and their parents as more concerned.

5. Teachers in high- and low-performing schools did not perceive different amounts of help available to them in handling problems.

6. More total activity takes place in reading classes in high-performing schools than in low-performing schools.

7. Children in reading classes in high-performing schools engaged in more silent reading while children in low-performing schools engaged in more oral reading.

8. Reading teachers' evaluations of classroom reading programs were on the average more favorable in high-performing schools. Reading teachers rated teachers in high-performing schools more favorably than they did teachers in low-performing schools in using appropriate material, extending reading into other areas, asking children to read with a purpose, and using informal diagnosis.

9. In grades one to three, teachers in high-performing schools gave more positive and less negative reinforcement than did teachers in low-performing schools.

10. In grades four to six, teachers in low-performing schools gave more reinforcement. In general, however, teachers in high-performing schools gave positive reinforcement more than negative reinforcement.

11. On selected items related to open education, high-performing schools appeared more open than low-performing schools.

12. On several measures, differences between high- and low-performing schools seemed more pronounced in grades one to three than in grades four to six.

13. Of three instruments tested, two significantly differentiated high- and low-performing schools arranged on a rank-order basis. Thus, not only did individual items differentiate between schools, but total rankings on the two instruments distinguished successfully between the two groups of schools.

14. Principals in high-performing schools generally saw their personnel as more competent than did principals in low-performing schools.

15. Principals in high-performing schools saw themselves as having better rapport with teachers, parents, and pupils than did principals in low-performing schools. Principals in low-performing schools reported better rapport with their school boards, however.

16. Items on physical space and facilities generally did not differentiate between high- and low-performing schools.

17. In general, reasonably close relationships were found between the ratings of different observers.

REFERENCES

1. Bureau of School Programs Evaluation. Variables Related to Student Performance and Resource Allocation Decisions at the School District Level. Albany, New York: State Education Department. June 1972.

2. _____. What Research Says About Improving Student Performance. Albany, New York: State Education Department. March 1973.

3. _____. Performance Indicators in Education Technical Manual. Albany, New York: State Education Department. Spring 1974.

4. Bottenberg, Robert A., and Ward, Joe H., Jr. Applied Multiple Linear Regression. Technical Documentary Report PRL-TDR-63-6. 6507th Personnel Research Laboratory, Lackland Air Force Base, Texas. 1963.

5. Coleman, James S., et al. Equality of Educational Opportunity. Washington, D. C.: National Center for Educational Statistics, U.S. Government Printing Office. 1966.

6. Mayeske, George W., et al. A Study of Our Nation 's Schools. Washington, D.C.: U.S. Department of Health, Education, and Welfare. 1972.

7. Spath, Guy D., and Pillsworth, Philip J. Using Interaction and Curvilinear Terms to Improve New York State Performance Models. Albany, New York: State Education Department, Bureau of School Programs Evaluation. April 1973.

8. Spath, Guy D. Does Attendance Rate Make a Difference on Reading and Arithmetic Achievement? Albany, New York: State Education Department, Bureau of School Programs Evaluation. August 1974.

9. Spath, Guy D., and DiGiovanni, Gennaro. District Wealth and Student Achievement: Does a Relationship Exist? Albany, New York: State Education Department, Bureau of School Programs Evaluation. November 1974.

10. Spath, Guy D. Identifying Expenditure Variables Which Relate to Student Performance at the District Level. Albany, New York: State Education Department, Bureau of School Programs Evaluation. January 1975.

11. Spath, Guy D. Does Student Mobility Affect School District Results on Reading and Math Achievement. Albany, New York: State Education Department, Bureau of School Programs Evaluation. January 1975.

12. DiGiovanni, Gennaro. The Relationship Between Percent of Certified Teachers and Student Achievement in New York School Districts. Albany, New York: State Education Department, Bureau of School Programs Evaluation. March 1975 .

13. DiGiovanni, Gennaro. Personal and Professional Teacher Characteristics and Student Achievement in New York State. Albany, New York: State Education Department, Bureau of School Programs Evaluation. April 1975.

14. DiGiovanni, Gennaro. Teacher Salaries and Student Achievement in New York State. Albany, New York: State Education Department, Bureau of School Programs Evaluation. April 1975.

15. Nichols, Robert C., and Swanson, Austin D. Confidential Report on A Study of School Outliers in New York State. Buffalo: Faculty of Educational Studies, State University of New York at Buffalo. August 1974.

16. Swanson, Austin D. Supplementary Report on A Study of School Outliers in New York State. Buffalo: Faculty of Educational Studies, State University of New York at Buffalo. March 1975.

17. Division of Education Evaluation. Reading Achievement Related to Educational and Environmental Conditions in 12 New York City Elementary Schools. Albany, New York: State Education Department. March 1974.

18. Clark, Richard M. An Ecological Study of Teachers and Administrators in Fourteen Schools. Albany: Capital Area Research and Field Services. March 1975.

19. Soar, S., and Soar, R.M. "An empirical analysis of selected early childhood programs." Early Childhood Education. National Society for the Study of Education. Part 2. Pages 229-260. 1972.

20. Quirk, T.J., Weinberg, S.F., and Nalin, K.B. The Development of a Student Observation Instrument for Reading Instruction: Project Report. Princeton, New Jersey: Educational Testing Service. 1973.

21. Weinberg, S., Nalin, K., and Quirk, T.J. The Classroom Behavior of Students During Compensatory Reading Instruction. Prepared for American Educational Research Association annual meeting. Chicago. 1974.

22. Walberg, H.J., and Thomas, S.C. Characteristics of Open Education: Toward an Operational Definition. Newton, Massachusetts: Education Development Center. 1971.